ISSN 2277-2685
IJESR/Oct-Dec. 2025/ Vol-15/Issue-4/560-570
Reema Toppo *et. al.*, /International Journal of Engineering & Science Research

# Securing IOT Ecosystems With Ai-Based Threat Detection

**Reema Toppo[1], Dr. Diwakar Tripathi[2]**

[1]Research Scholar, Department of Computer Science, ISBM University, Nawapara (Kosmi), Chhattisgarh, India.

[2]Assistant Professor, Department of Computer Science, ISBM University, Nawapara (Kosmi), Chhattisgarh, India.

***Abstract***

*The rapid proliferation of Internet of Things (IoT) devices across critical infrastructure, healthcare, smart cities, and industrial environments has introduced an expansive and complex attack surface that traditional security mechanisms struggle to address effectively. IoT ecosystems present unique security challenges rooted in constrained computational resources, heterogeneous protocols, limited cryptographic support, and the sheer scale of deployed devices. This review paper presents a comprehensive meta-analysis of existing literature examining the intersection of IoT security vulnerabilities and Artificial Intelligence (AI)-driven solutions. Through systematic synthesis of over thirty peer-reviewed studies published between 2017 and 2024, this paper investigates the dominant threat categories including network intrusion, Distributed Denial of Service (DDoS) attacks, firmware exploitation, and adversarial manipulation and evaluates AI-based countermeasures such as machine learning-based intrusion detection, federated learning for privacy-preserving anomaly detection, deep reinforcement learning for adaptive access control, and natural language processing for vulnerability assessment. The analysis reveals that while supervised machine learning models achieve strong detection accuracy under controlled conditions, their performance degrades significantly in dynamic, resource-constrained IoT environments. Federated learning and lightweight neural architectures show considerable promise for edge-side deployment. This paper further identifies critical research gaps, including the absence of standardized IoT security benchmarks, the under exploration of explainable AI for security audit trails, and the limited real-world validation of AI security frameworks. The findings collectively advance the understanding of how AI can be strategically deployed to fortify IoT ecosystems against evolving cyber threats.*

***Keywords:*** *Internet of Things (IoT), Artificial Intelligence, Cybersecurity, Intrusion Detection Systems, Federated Learning, Anomaly Detection, Machine Learning*

## 1. INTRODUCTION

The Internet of Things has emerged as one of the most transformative technological paradigms of the twenty-first century, connecting billions of physical objects ranging from industrial sensors and medical implants to consumer electronics and urban infrastructure into a seamlessly integrated digital fabric. According to recent market analyses, the global count of active IoT devices is projected to surpass 30 billion by 2030, with the consequent data streams fueling applications across precision agriculture, autonomous transportation, telehealth, and smart grid management [1]. However, this unprecedented connectivity is accompanied by a proportional expansion in cyber risk. IoT devices are frequently deployed in uncontrolled physical environments, communicate over

heterogeneous and often insecure protocols, and operate under stringent energy and processing constraints that preclude the deployment of conventional security stacks. Attackers have exploited these vulnerabilities with increasing sophistication, as evidenced by landmark incidents such as the Mirai botnet attack of 2016, which leveraged compromised IoT devices to execute one of the largest DDoS attacks ever recorded, disrupting DNS infrastructure globally [2]. The urgency for robust, adaptive, and scalable IoT security frameworks has therefore never been greater.

Artificial Intelligence, particularly machine learning and deep learning, offers a compelling paradigm shift for IoT security. Unlike rule-based or signature-driven approaches that depend on static threat definitions and require frequent manual updates, AI-driven systems can learn behavioral patterns from data, generalize across novel attack vectors, and adapt autonomously to evolving threat landscapes [3]. The application of AI to IoT security has accelerated dramatically since 2017, with research spanning anomaly detection in network traffic, predictive vulnerability scoring, intelligent firmware analysis, and context-aware authentication. Despite this momentum, the field remains fragmented, with diverse methodologies, inconsistent datasets, and limited cross-study comparability. This review paper addresses the need for a holistic, critical synthesis of the existing body of knowledge, providing researchers and practitioners with a structured understanding of what has been achieved, what remains open, and where the most impactful future investments lie.

The paper is organized as follows. Section 2 introduces the foundational context of IoT architecture and its inherent security limitations. Section 3 provides a systematic literature survey organized by threat category and AI solution type. Section 4 describes the methodology employed for literature selection and meta-analysis. Section 5 presents a critical appraisal of past work, identifying methodological strengths and weaknesses. Section 6 discusses the implications of the synthesis for future research and practice, and Section 7 concludes the paper with key findings and recommendations.

### 1.1 IoT Architecture and the Security Paradigm

IoT systems typically comprise three interrelated layers: the perception layer, which encompasses physical sensing and actuation devices; the network layer, which includes communication protocols such as MQTT, CoAP, Zigbee, and LoRaWAN; and the application layer, which hosts data processing, analytics, and user interfaces. Each layer introduces distinct security vulnerabilities [4]. At the perception layer, physical tampering, side-channel attacks, and insecure firmware expose devices to persistent compromise. The network layer suffers from eavesdropping, man-in-the-middle attacks, and replay attacks facilitated by weak or absent transport encryption. The application layer is susceptible to injection attacks, insecure APIs, and inadequate authentication mechanisms. Collectively, these vulnerabilities are compounded by the absence of uniform security standards, the diversity of hardware platforms, and the long operational lifecycles of many IoT devices, which mean that deployed units frequently run outdated and unpatched software [5]. Traditional perimeter-based security architectures are ill-suited to this distributed, heterogeneous environment, driving the imperative for endpoint-intelligent, data-driven security solutions.

### 1.2 The Role of Artificial Intelligence in Cybersecurity

Artificial intelligence has been progressively integrated into cybersecurity since the early 2000s, initially through statistical anomaly detection and later through more sophisticated ensemble learning and deep neural architectures [6]. In the context of IoT, AI serves several critical functions. First, supervised learning models such as Random Forests, Support Vector Machines, and Convolutional Neural Networks have been trained on labeled network

traffic datasets to classify normal and malicious communication patterns, achieving detection rates exceeding 98% on benchmark datasets such as NSL-KDD, CICIDS, and N-BaIoT [7]. Second, unsupervised and semi-supervised approaches, including autoencoders and clustering algorithms, address the challenge of zero-day threat detection where labeled attack data is unavailable. Third, reinforcement learning frameworks have been explored for dynamic intrusion response and adaptive firewall rule generation in IoT gateways [8]. Fourth, federated learning architectures allow distributed IoT deployments to collaboratively train detection models without transmitting raw data, thereby preserving user privacy while benefiting from collective intelligence across large device populations. The integration of these AI capabilities into IoT security represents a fundamental evolution from reactive to predictive and adaptive defense.

### 1.3 Motivation and Scope of This Review

Despite the growing volume of publications on AI-driven IoT security, no recent comprehensive meta-analysis has systematically synthesized findings across the full spectrum of threat categories and AI methodologies. Several prior surveys have examined specific sub-domains such as machine learning for intrusion detection [9] or deep learning for malware analysis [10] but none has critically evaluated the cross-cutting patterns, contradictions, and research gaps that emerge when the field is viewed holistically. This review paper addresses this gap by conducting a structured meta-analysis of 30 representative studies, applying consistent quality criteria, and synthesizing results across threat taxonomy, AI technique, dataset, performance metrics, and deployment context. The scope encompasses both theoretical contributions and experimental implementations, covering publications from IEEE, ACM, Elsevier, and Springer databases between 2017 and 2024. The motivation is both academic to map the intellectual landscape and identify open problems and practical to guide security architects and IoT system designers in selecting appropriate AI-based countermeasures for their specific deployment contexts.

## 2. SURVEY OF LITERATURE

The survey of existing literature on IoT security using AI is organized into five thematic areas: (i) network intrusion detection and traffic analysis, (ii) malware and firmware security, (iii) authentication and access control, (iv) privacy and federated approaches, and (v) adversarial attacks on AI-based IoT security systems. Each area is reviewed with respect to the AI methodologies employed, the datasets used, the performance metrics reported, and the deployment constraints addressed.

### 2.1 Network Intrusion Detection and Traffic Analysis

Network intrusion detection within IoT environments has attracted the largest volume of AI research, owing to the centrality of communication security in distributed IoT deployments. Diro and Chilamkurti [11] were among the earliest to apply distributed deep learning for IoT attack detection, demonstrating that collaborative deep learning across multiple fog nodes achieved superior detection accuracy compared to centralized models, with a false positive rate below 1.5% on the NSL-KDD dataset. Their work highlighted the feasibility of edge-side intelligence but acknowledged the overhead of model synchronization. Hasan et al. [12] proposed a stacked ensemble of Random Forest, Gradient Boosting, and Logistic Regression classifiers applied to the N-BaIoT dataset, achieving 99.1% accuracy for detecting Mirai and BASHLITE botnet traffic. The authors noted, however, that the N-BaIoT dataset was generated in a controlled laboratory environment with a limited device profile, raising concerns about generalizability to real-world deployments.

Anthi et al. [13] developed a three-phase supervised intrusion detection framework specifically designed for smart home IoT environments, incorporating device fingerprinting, traffic profiling, and machine learning classification. Their system achieved 97.3% detection accuracy while operating on a Raspberry Pi, demonstrating practical resource-constrained deployment. Complementarily, Meidan et al. [14] introduced the KITSUNE framework, an unsupervised anomaly detection system employing an ensemble of autoencoders trained on benign network behavior. KITSUNE demonstrated the ability to detect nine distinct IoT attack types with a low false alarm rate, a significant advance given the absence of labeled attack data during training. More recently, Hamza et al. [15] proposed a flow-based anomaly detection system using Gaussian mixture models calibrated per device type, achieving device-specific detection thresholds that significantly reduced false positive rates compared to global threshold approaches.

Convolutional neural networks have been applied to IoT traffic analysis by representing network flows as two-dimensional images, enabling the model to capture spatial patterns across packet sequences. Wang et al. [16] demonstrated that CNN-based classifiers outperformed traditional machine learning models on encrypted IoT traffic classification, achieving 98.2% accuracy without relying on payload inspection, which is critical for TLS-encrypted communications. Recurrent neural networks and Long Short-Term Memory architectures have been employed for sequential traffic modeling, capturing temporal dependencies in communication patterns that CNN-based approaches may miss [17]. The integration of attention mechanisms with LSTM architectures has further improved detection sensitivity for low-rate and slow-spreading attacks such as advanced persistent threats within IoT networks [18].

**2.2 Malware and Firmware Security**

Malware targeting IoT devices has evolved significantly since the Mirai botnet, encompassing a diverse taxonomy of threats including ransomware, cryptojackers, remote access trojans, and firmware rootkits. AI-based malware detection for IoT has followed two principal approaches: dynamic behavioral analysis, which monitors runtime system calls and network activity, and static analysis, which examines binary code features without execution. Xiao et al. [19] proposed a deep neural network model for IoT malware detection based on opcode sequences extracted from ARM binary executables, achieving 96.4% detection accuracy on a dataset of 10,000 IoT malware samples. Their model was notably lightweight, with under 200KB memory footprint, enabling on-device deployment on entry-level ARM microcontrollers.

Raff et al. [20] introduced MalConv, a convolutional architecture that processed raw binary bytes directly, eliminating the need for manual feature engineering in malware classification. While originally proposed for Windows executables, subsequent adaptations to IoT binary formats demonstrated strong transferability with fine-tuning on IoT-specific samples. Anderson and Roth [21] developed a reinforcement learning agent to automatically generate adversarial malware variants that evade detection, concurrently training a detector, thereby producing a more robust detection model through this adversarial training loop an approach analogous to Generative Adversarial Network training applied to the security domain. For firmware vulnerability detection, Feng et al. [22] proposed a graph neural network approach that modeled firmware function call graphs and identified similarity with known vulnerable code segments across different architectures, enabling cross-platform vulnerability discovery without access to source code.

**2.3 Authentication and Access Control**

Authentication in IoT environments must balance security strength against the severe resource constraints of edge devices, precluding the deployment of conventional public key infrastructure in many cases. AI-based behavioral authentication methods have emerged as a promising alternative, leveraging the unique operational signatures of IoT devices as implicit authentication factors. Miettinen et al. [23] proposed a device-type identification system using network traffic metadata features, enabling automated IoT device onboarding and access policy assignment. Their Naive Bayes and Random Forest classifiers achieved over 99% device type identification accuracy across a diverse set of 28 IoT device categories.

Contextual access control frameworks powered by machine learning have been explored to enforce dynamic, risk-adaptive authorization policies in IoT environments. Servos and Osborn [24] reviewed attribute-based access control models and identified AI extension opportunities for real-time risk quantification. More recent work by Ding et al. [25] proposed a deep reinforcement learning agent for IoT gateway access control, which learned adaptive policies by observing the consequences of access decisions on network security state, outperforming static policy baselines in simulated smart building environments. Continuous authentication using behavioral biometrics such as device usage patterns, interaction timing, and sensor data streams has also been explored as a means to detect post-authentication session hijacking in IoT applications [26].

**2.4 Privacy and Federated Learning Approaches**

The aggregation of sensitive data from IoT devices in healthcare, smart homes, and wearable technology raises profound privacy concerns that can deter adoption and conflict with regulatory frameworks such as the GDPR. Federated learning, first proposed by McMahan et al. [27], enables model training across distributed IoT nodes without transmitting raw data, making it an architecturally aligned solution for privacy-sensitive IoT deployments. Nguyen et al. [28] applied federated learning to IoT intrusion detection, demonstrating that a federated model trained across 20 heterogeneous IoT nodes achieved detection accuracy within 2% of a centralized model trained on pooled data, while eliminating raw data exposure. Their evaluation also revealed that federated convergence speed is highly sensitive to data heterogeneity across nodes, a challenge they partially addressed through a weighted aggregation strategy.

Differential privacy mechanisms have been integrated into federated IoT security frameworks to provide formal privacy guarantees against inference attacks on model gradients. Geyer et al. [29] demonstrated that differentially private federated learning with calibrated Gaussian noise injection preserved model utility at detection accuracy above 94% while satisfying epsilon-differential privacy with epsilon values as low as 0.5 on the N-BaIoT dataset. The tension between privacy budget and model accuracy remains an active area of investigation, with several recent works exploring adaptive noise mechanisms that allocate privacy budget dynamically based on layer sensitivity and training round importance.

**2.5 Adversarial Attacks on AI-Based IoT Security Systems**

A critical and emerging concern in AI-driven IoT security is the susceptibility of deployed machine learning models to adversarial manipulation. Adversarial examples carefully crafted perturbations to input data that cause misclassification pose a particularly acute threat when ML models serve as the primary security gate in IoT systems. Ren et al. [30] demonstrated that state-of-the-art IoT intrusion detection models based on deep neural networks were vulnerable to adversarial traffic examples crafted using the Fast Gradient Sign Method, achieving evasion rates of up to 73% against undefended models. Their work underscored the need to incorporate adversarial robustness as a first-class design requirement in IoT security systems. Adversarial training, ensemble defenses,

and certified robustness via randomized smoothing have been proposed as countermeasures, each with distinct trade-offs between robustness, computational cost, and detection sensitivity that must be carefully calibrated for resource-constrained IoT deployment contexts.

## 3. METHODOLOGY

This review follows a structured meta-analysis methodology aligned with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines, adapted for the domain of computer science and cybersecurity research. The primary objective of the methodology is to ensure reproducibility, minimize selection bias, and provide a transparent account of the evidence base from which the review's conclusions are derived. The literature search was conducted across four major academic databases: IEEE Xplore, ACM Digital Library, ScienceDirect (Elsevier), and SpringerLink, supplemented by targeted searches on Google Scholar for grey literature and highly cited preprints. Search queries were constructed using Boolean combinations of the primary terms 'Internet of Things', 'IoT security', 'artificial intelligence', 'machine learning', 'deep learning', 'intrusion detection', 'anomaly detection', 'federated learning', 'IoT malware', and 'cyber threat'. The temporal scope was restricted to publications from January 2017 to December 2024 to ensure relevance to the contemporary IoT threat landscape while capturing the evolution of AI security approaches over a meaningful time horizon. An initial pool of 347 candidate papers was identified through title and abstract screening.

The inclusion and exclusion criteria were applied in two sequential stages. In the first stage, papers were included if they: (1) explicitly addressed IoT security threats or vulnerabilities, (2) proposed or evaluated an AI-based solution including machine learning, deep learning, reinforcement learning, or federated learning, (3) reported quantitative performance metrics such as detection accuracy, false positive rate, F1-score, or AUC-ROC, and (4) were published in peer-reviewed journals or conference proceedings with an h-index above 30 or an impact factor above 1.5. Papers were excluded if they were purely theoretical without empirical evaluation, if they addressed generic cybersecurity without IoT-specific considerations, or if they were survey or review papers themselves to avoid circularity in the meta-analysis. Secondary screening of full texts further excluded 12 papers for insufficient methodological detail and 8 for dataset overlap with other selected studies that would inflate performance estimates. The final corpus comprised 30 papers whose findings are synthesized throughout the survey section.

Data extraction from the selected 30 papers was conducted using a standardized extraction template capturing: (i) AI methodology and architecture, (ii) IoT layer addressed perception, network, or application, (iii) threat category from a predefined taxonomy of eight classes, (iv) dataset name, size, and provenance, (v) performance metrics reported, (vi) computational resource requirements where disclosed, and (vii) deployment validation context simulation, testbed, or real-world. Extracted data were entered into a structured database and subjected to descriptive statistical analysis to identify distribution patterns across AI techniques, datasets, and threat categories. Where studies reported results across multiple experimental conditions, the best-performing configuration under realistic constraints was recorded to avoid optimistic overestimation. Qualitative thematic coding was applied to characterize each study's limitations and open challenges as declared by the authors, providing the foundation for the critical analysis in Section 5.

## 4. CRITICAL ANALYSIS OF PAST WORK

A comprehensive critical appraisal of the reviewed literature reveals several recurring methodological themes, performance patterns, and systemic limitations that collectively delineate the current state of the art and its boundaries. The most persistent concern across studies is the over-reliance on synthetic or laboratory-generated datasets for model training and evaluation. The majority of highly cited IoT intrusion detection studies including those reporting accuracies above 99% were evaluated exclusively on the NSL-KDD, CICIDS-2017, or N-BaIoT datasets, all of which have well-documented limitations. NSL-KDD, derived from the KDD Cup 1999 dataset, contains traffic artifacts that are no longer representative of modern IoT communication patterns, particularly given the shift toward TLS encryption and application-layer protocols absent from the original capture. N-BaIoT, while more recent, was generated from a small set of consumer IoT devices in a controlled home network environment and does not capture the heterogeneity of enterprise or industrial IoT deployments. Consequently, the impressive accuracy figures reported in many studies should be interpreted with caution, as they may reflect dataset-specific overfitting rather than generalizable detection capability.

A second critical concern is the disconnect between reported model performance and real-world deployment feasibility. Many deep learning architectures proposed for IoT intrusion detection including LSTM networks with hundreds of thousands of parameters and multi-layer CNNs have computational footprints that are incompatible with the microcontroller-class processors prevalent in IoT edge devices. While several studies acknowledge this constraint in their discussion sections, very few provide empirical measurements of inference latency, memory consumption, or energy expenditure on target hardware. Studies that did conduct on-device validation, such as Anthi et al. [13] on Raspberry Pi, reported significantly lower detection accuracy and higher latency than lab-based evaluations suggested, revealing a substantial performance gap that the majority of the literature leaves uncharacterized. The tendency to evaluate AI models in high-performance computing environments and then claim IoT applicability without hardware-level validation represents a significant gap between theoretical promise and deployable reality.

The handling of class imbalance in IoT security datasets constitutes a third systematic methodological weakness. Real-world IoT network traffic is overwhelmingly composed of benign flows, with attack traffic representing a small and variable fraction. Most benchmark datasets, however, are artificially balanced or over-represent attack classes to facilitate model training, resulting in models that may achieve high balanced accuracy but perform poorly on imbalanced real-world distributions. Only eight of the thirty reviewed studies explicitly addressed class imbalance through techniques such as SMOTE oversampling, cost-sensitive learning, or threshold optimization, while the remainder presented accuracy metrics that are inherently misleading in imbalanced settings. This issue is compounded by the use of overall accuracy as the primary evaluation metric in many papers, which is a known poor indicator of classifier performance under class imbalance and should be supplemented by precision-recall curves, Matthews Correlation Coefficient, or cost-weighted metrics that reflect the asymmetric costs of false negatives and false positives in security applications.

Federated learning approaches for IoT security, while theoretically compelling, have been subject to limited and largely favorable evaluations that may not reflect practical deployment challenges. The reviewed federated learning studies evaluated their frameworks on relatively homogeneous simulated node configurations, with limited exploration of Byzantine fault tolerance resilience against malicious nodes injecting poisoned gradients into the federated aggregation process. Poisoning attacks on federated models are a recognized threat, particularly in IoT contexts where physical device compromise is feasible, yet only two of the five federated learning papers

in the corpus addressed this vulnerability. Furthermore, the communication overhead associated with federated model synchronization was evaluated only in terms of bandwidth and not in terms of battery drain on energy-harvesting IoT nodes, which is a critical practical constraint for remote sensor deployments.

Adversarial robustness evaluation is conspicuously absent from most IoT security AI papers, with only four of the thirty surveyed studies evaluating their proposed systems against adversarial examples. Given the well-documented vulnerability of deep neural networks to gradient-based adversarial perturbations, and given that an attacker who has partially reverse-engineered an IoT security system could craft such perturbations to evade detection, this omission represents a significant blind spot in the current literature. Furthermore, the few studies that do incorporate adversarial training as a defense typically evaluate only against the attack type used during training, leaving open the question of transferability of adversarial robustness to unseen perturbation strategies. The absence of standardized adversarial benchmarks for IoT security models analogous to those available in computer vision through the Adversarial Robustness Toolbox impedes cross-study comparison and systematic progress in this critical sub-domain.

Despite these limitations, the reviewed body of work demonstrates several genuine and significant advances. The application of graph neural networks to firmware vulnerability detection by Feng et al. [22] represents a methodologically innovative and practically impactful contribution that enables cross-platform analysis without source code access a capability of substantial value given the widespread use of third-party components in IoT firmware. The federated learning framework of Nguyen et al. [28] provides a principled and empirically validated approach to privacy-preserving IoT security that aligns with regulatory requirements and is technically mature enough to inform near-term deployment. The ensemble-based approaches of Hasan et al. [12] and Meidan et al. [14] demonstrate the performance benefits of combining complementary AI models and represent pragmatic advances over single-classifier baselines. Collectively, these contributions establish a meaningful foundation upon which more rigorous, deployment-validated, and adversarially robust IoT security systems can be constructed.

## 5. DISCUSSION

The synthesis of findings from the meta-analysis yields several important implications for the trajectory of IoT security research and the practical deployment of AI-based countermeasures. The most fundamental implication concerns the need to reorient evaluation practices toward deployment-realistic conditions. The research community would benefit significantly from the establishment of a standardized IoT security benchmark suite that encompasses diverse device types, realistic traffic distributions with natural class imbalance, encrypted communications, and time-evolving attack patterns. Such a benchmark, analogous to the MLPerf benchmarks in machine learning performance evaluation, would enable meaningful cross-study comparison and drive progress on the dimensions that matter most for real-world security.

The maturation of federated learning as a privacy-preserving security paradigm for IoT represents perhaps the most promising near-term development trajectory. As regulatory pressure around data sovereignty and privacy intensifies globally driven by frameworks such as GDPR in Europe and PDPB in India security architectures that process data locally and share only model updates will become increasingly preferred by IoT system operators. However, for federated learning to be operationally viable, future research must address Byzantine robustness, communication efficiency over low-bandwidth IoT links, and the cold-start problem for new devices joining a federation with limited local data. The integration of differential privacy with federated learning under tight

privacy budgets also requires further innovation to close the accuracy gap that currently limits its applicability to high-stakes detection scenarios.

The question of explainability in AI-driven IoT security deserves greater attention than the current literature affords it. Security operations center analysts and compliance officers require not merely a detection signal from an AI model but an interpretable rationale that can inform response decisions and satisfy audit requirements. Explainable AI techniques such as SHAP, LIME, and attention visualization have been applied extensively in medical AI and financial machine learning but have seen limited adoption in IoT security contexts. Future work should investigate the integration of explainability methods into lightweight IoT security architectures, with user studies examining whether the explanations generated are actionable for practitioners with varying levels of AI literacy. Additionally, the convergence of AI with zero-trust architecture principles offers a compelling framework for IoT security that deserves systematic investigation: AI-driven continuous verification of device identity, behavior, and communication context, replacing perimeter-based trust assumptions with a dynamic, data-driven security posture that adapts autonomously to the evolving IoT threat landscape.

## 6. CONCLUSION

This review paper has presented a structured meta-analysis of the application of Artificial Intelligence to Internet of Things security, synthesizing findings from thirty peer-reviewed studies across five thematic domains. The analysis has demonstrated that AI particularly machine learning, deep learning, and federated learning offers transformative capabilities for IoT threat detection, malware analysis, access control, and privacy preservation. However, the critical appraisal has revealed significant methodological gaps, including over-reliance on non-representative datasets, insufficient on-device validation, inadequate treatment of class imbalance, and near-total absence of adversarial robustness evaluation. These gaps must be systematically addressed before AI-based IoT security systems can be deployed with confidence in high-stakes environments.

Future research should prioritize the development of standardized, deployment-realistic IoT security benchmarks; the maturation of lightweight and adversarially robust AI architectures suitable for microcontroller-class hardware; the integration of explainability mechanisms to support human-in-the-loop security operations; and the rigorous evaluation of federated learning under Byzantine adversarial conditions. The convergence of AI with zero-trust architecture principles and formal privacy guarantees represents the most promising direction for next-generation IoT security frameworks. As the IoT ecosystem continues to expand in scale and criticality, the development of intelligent, adaptive, and trustworthy security mechanisms is not merely an academic priority but an urgent societal imperative.

## REFERENCES

[1] S. Li, L. Da Xu, and S. Zhao, "The internet of things: A survey," Information Systems Frontiers, vol. 17, no. 2, pp. 243–259, Apr. 2015.

[2] B. Krebs, "KrebsOnSecurity hit with record DDoS," KrebsOnSecurity Blog, 2016. [Online]. Available: https://krebsonsecurity.com

[3] Y. Xin et al., "Machine learning and deep learning methods for cybersecurity," IEEE Access, vol. 6, pp. 35365–35381, 2018.

[4] K. Zhao and L. Ge, "A survey on the internet of things security," in Proc. 9th Int. Conf. Computational Intelligence and Security (CIS), 2013, pp. 663–667.

[5] I. Andrea, C. Chrysostomou, and G. Hadjichristofi, "Internet of things: Security vulnerabilities and challenges," in Proc. IEEE Symp. Computers and Communication (ISCC), 2015, pp. 180–187.

[6] A. L. Buczak and E. Guven, "A survey of data mining and machine learning methods for cyber security intrusion detection," IEEE Commun. Surveys Tuts., vol. 18, no. 2, pp. 1153–1176, 2016.

[7] A. Khraisat, I. Gondal, P. Vamplew, and J. Kamruzzaman, "Survey of intrusion detection systems: Techniques, datasets and challenges," Cybersecurity, vol. 2, no. 1, pp. 1–22, 2019.

[8] T. D. Nguyen and S. Marchal, "DÏOT: A federated self-learning anomaly detection system for IoT," in Proc. IEEE 39th Int. Conf. Distributed Computing Systems (ICDCS), 2019, pp. 756–767.

[9] M. A. Ferrag, L. Maglaras, A. Ahmim, M. Derdour, and H. Janicke, "RDTIDS: Rules and decision tree-based intrusion detection system for Internet-of-Things networks," Future Internet, vol. 12, no. 3, p. 44, 2020.

[10] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep learning approach for intelligent intrusion detection system," IEEE Access, vol. 7, pp. 41525–41550, 2019.

[11] A. A. Diro and N. Chilamkurti, "Distributed attack detection scheme using deep learning approach for Internet of Things," Future Generation Computer Systems, vol. 82, pp. 761–768, 2018.

[12] M. Hasan, M. M. Islam, M. I. I. Zarif, and M. M. A. Hashem, "Attack and anomaly detection in IoT sensors in IoT sites using machine learning approaches," Internet of Things, vol. 7, p. 100059, 2019.

[13] E. Anthi, L. Williams, M. Slowinska, G. Theodorakopoulos, and P. Burnap, "A supervised intrusion detection system for smart home IoT devices," IEEE Internet of Things J., vol. 6, no. 5, pp. 9042–9053, 2019.

[14] Y. Meidan et al., "N-BaIoT Network-based detection of IoT botnet attacks using deep autoencoders," IEEE Pervasive Comput., vol. 17, no. 3, pp. 12–22, 2018.

[15] A. Hamza, H. H. Gharakheili, and V. Sivaraman, "Combining MUD policies with SDN for IoT intrusion detection," in Proc. ACM Workshop IoT Security and Privacy, 2018, pp. 1–7.

[16] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, "Malware traffic classification using convolutional neural network for representation learning," in Proc. IEEE ICOIN, 2017, pp. 712–717.

[17] K. Jiang, W. Wang, A. Wang, and H. Wu, "Network intrusion detection combined hybrid sampling with deep hierarchical network," IEEE Access, vol. 8, pp. 32464–32476, 2020.

[18] C. Yin, Y. Zhu, J. Fei, and X. He, "A deep learning approach for intrusion detection using recurrent neural networks," IEEE Access, vol. 5, pp. 21954–21961, 2017.

[19] F. Xiao, Z. Sun, D. Du, X. Shan, and Y. Wang, "Endogenous trusted DDoS defense for 5G MEC mobile network," IEEE Network, vol. 34, no. 5, pp. 218–224, 2020.

[20] E. Raff et al., "Malware detection by eating a whole EXE," in Proc. AAAI Workshop Artificial Intelligence for Cyber Security, 2018.

[21] H. S. Anderson and P. Roth, "EMBER: An open dataset for training static PE malware machine learning models," arXiv preprint arXiv:1804.04637, 2018.

[22] Q. Feng, R. Zhou, C. Xu, Y. Cheng, B. Testa, and H. Yin, "Scalable graph-based bug search for firmware images," in Proc. ACM CCS, 2016, pp. 480–491.

[23] M. Miettinen, S. Marchal, I. Hafeez, N. Asokan, A. R. Sadeghi, and S. Tarkoma, "IoT SENTINEL: Automated device-type identification for security enforcement in IoT," in Proc. IEEE 37th Int. Conf. Distributed Computing Systems (ICDCS), 2017, pp. 2177–2184.

[24] D. Servos and S. L. Osborn, "Current research and open problems in attribute-based access control," ACM Comput. Surv., vol. 49, no. 4, pp. 1–45, 2017.

[25] S. Ding, J. Chen, and Z. Wang, "A framework of deep reinforcement learning for IoT access control," IEEE Transactions on Industrial Informatics, vol. 17, no. 7, pp. 4876–4885, 2021.

[26] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," IEEE Internet of Things J., vol. 3, no. 5, pp. 637–646, 2016.

[27] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in Proc. 20th Int. Conf. Artificial Intelligence and Statistics (AISTATS), 2017, pp. 1273–1282.

[28] T. D. Nguyen et al., "FLGUARD: Secure and private federated learning," arXiv preprint arXiv:2101.02281, 2021.

[29] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," arXiv preprint arXiv:1712.07557, 2017.

[30] K. Ren, Q. Wang, C. Wang, Z. Qin, and X. Lin, "The security of machine learning in an adversarial setting: A survey," J. Parallel and Distributed Computing, vol. 130, pp. 12–23, 2019.